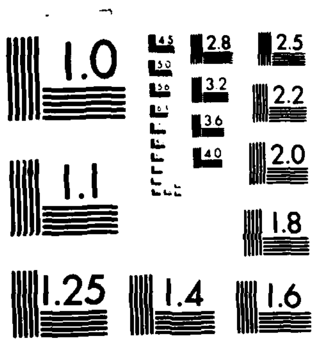
AD-A129 281 COMPARATIVE NOTCHED BOX PLOTS(U) PENNSYLVANIA STATE 1/1
UNIV UNIVERSITY PARK DEPT OF STATISTICS
T P HETTMANSPERGER JUN 83 TR-43 N00014-80-C-0741

UNCLASSIFIED F/G 12/1 NL

END
DATE
FILMED
7 83
DTIC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0

12

DEPARTMENT OF STATISTICS

The Pennsylvania State University

University Park, Penn., 16802, U.S.A.

TECHNICAL REPORTS AND PREPRINTS

Number 43: June 1983

COMPARATIVE NOTCHED BOX PLOTS

by

Thomas P. Hettmansperger

The Pennsylvania State University, University Park

Research partially supported by ONR Contract N00014-80-C-0741

- a -

COMPARATIVE NOTCHED BOX PLOTS

Thomas P. Hettmansperger
The Pennsylvania State University

ABSTRACT

Comparative notched box plots are developed to provide confidence intervals and hypothesis tests for the two sample location model. The notches are confidence intervals derived from the sign test. Rules are given for assigning confidence coefficients to the notches to yield a 95 percent confidence interval and 5 percent two sided test for the difference in locations. The test that rejects no location difference when the "sign" notches are disjoint is shown to be Mood's median test. Circumstances under which multiple comparisons can be carried out are also discussed.

Key Words: Sign test, Mood's median test, Confidence intervals

Research partially supported by ONR Contract N00014-80-C-0741

COMPARATIVE NOTCHED BOX PLOTS

by Thomas P. Hettmansperger
The Pennsylvania State University

1. INTRODUCTION AND SUMMARY

McGill, Tukey and Larsen (1978) discuss notched box plots as one way of displaying relevant sample information about a population. The box is determined by the sample quartiles (hinges) and locates the middle half of the population distribution. The whiskers are related to the interquartile range (hinge spread) and are useful in identifying stray observations. The notch portion of the plot is an approximate confidence interval for the population median.

When several samples are displayed together, it is natural to compare the notches and make rough significance statements about the two population medians under consideration. A two sided, two sample test consists in rejecting the null hypothesis of equal population medians when the notches are disjoint. As McGill et al. (1978, Section 7) point out, if 95% individual notches are selected then the significance level for the comparison is less than 1%, much too stringent for rough significance statements. Their solution is to construct the notches by taking the ends to be:

$$M \pm 1.7\ SE \qquad (1.1)$$

where M is the sample median and SE is a sample estimate of the asymptotic standard error of the sample median when sampling from a normal population. The factor 1.7 was "empirically chosen" to produce, on the average, a two sided 5% test that the two population medians are equal.

In this paper we consider notches based on pairs of ordered sample values. Just as the median occurs at the middle of the sample, the ends of the notch occur at a given depth from each end of the sample. For example if $n = 17$ then the fifth value in from each end provides a 95.1% confidence interval. See Noether (1976, Table E). Furthermore, this notch is not necessarily symmetric about the sample median as in the case of (1.1). Asymmetry in the notch reflects additional information in the sample.

Most texts on nonparametric statistics relate this notch (confidence interval), the median and the sign test. The confidence coefficient for the notch is determined by the binomial distribution (null distribution of the sign test). (Noether 1976, Chapter 12; Lehmann 1975, Chapter 4; Hollander and Wolfe 1972, Chapter 3.) Thus, exact rather than approximate confidence coefficients can be associated with these notches. The sign test, sample median and notch can be thought of as interrelated statistical procedures. In Section 3 we will show that comparing two "sign" notches is equivalent to constructing Mood's two sample median test and associated confidence interval. Thus, there is an interesting connection between the one sample "sign" procedures and the two sample Mood procedures.

Before turning to the proposed solutions we describe the one sample problem in the notation that will be used for the remainder of the paper.

Suppose $X_{(1)} \leq \ldots \leq X_{(n)}$ are the ordered values of a random sample from a continuous distribution with cdf $F(x - \theta_x)$. We will further *suppose* θ_x *is* the unique median. A $\gamma = 1 - \alpha$ confidence interval for θ_x is given by:

$$[L_x, U_x] = [X_{(d_x)}, X_{(n-d_x+1)}] \tag{1.2}$$

with $P(S < d_x) = \alpha/2$ where the distribution of S is binomial with parameters n and .5, i.e., b(n, .5). We will refer to d_x as the notch depth. Noether (1978, Table E) provides the d_x values; otherwise, they are easily found in a binomial table. The central limit theorem, with a continuity correction, yields

$$d_x \doteq \frac{n}{2} + .5 - Z_{\alpha/2} \frac{\sqrt{n}}{2} \tag{1.3}$$

where $Z_{\alpha/2}$ is the upper $\alpha/2$ percentage point of the standard normal distribution. Since we are dealing with a symmetric binomial distribution the approximation is adequate for sample sizes of at least 5. We generally take d_x to be the greatest integer in the right side of (1.3). This means the true confidence coefficient is bounded below by γ.

Given two ordered samples $X_{(1)} \leq \ldots \leq X_{(n_1)}$ and $Y_{(1)} \leq \ldots \leq Y_{(n_2)}$ from $F(x - \theta_x)$ and $F(y - \theta_y)$, respectively, we wish to pick the two notches $[L_x, U_x]$ and $[L_y, U_y]$ such that:

1. When the notches are disjoint we reject H_o: $\Delta = \theta_y - \theta_x = 0$ with significance level $\alpha_c \doteq .05$ where α_c is the specified comparison error rate and
2. the differences in the notches $[L_y - U_x, U_y - L_x]$, provide a $\gamma_c = 1 - \alpha_c \doteq .95$ confidence interval for $\Delta = \theta_y - \theta_x$.

The solution, which is developed in detail in Sections 2 and 3, is quite simple provided the sample size ratio is not more than 2 to 1. The confidence coefficients γ_x and γ_y should be chosen as close as possible to .84. Hence the two sample test and confidence interval for Δ are based on a pair of .84 "sign" notches. If a table is unavailable then from (1.3), with $Z_{\alpha/2} = 1.41$ corresponding to $\gamma_x = \gamma_y = .84$, take the notch depths to be the greatest integers in

$$\frac{n_i}{2} + .5 - 1.41\frac{\sqrt{n_i}}{2} \doteq \frac{n_i + 1}{2} - \sqrt{\frac{n_i}{2}}, \quad i = 1, 2. \tag{1.4}$$

When a significance level α_c other than .05 for the comparison is desired, the notch depths are taken to be the greatest integers in

$$\frac{n_i + 1}{2} - Z_{\alpha_c/2} \; \frac{1}{2}\sqrt{\frac{n_i}{2}}, \quad i = 1, 2. \tag{1.5}$$

The corresponding confidence coefficient is $\gamma_x = \gamma_y = 1 - 2\Phi(-Z_{\alpha_c/2}/\sqrt{2})$.

If the ratio of sample sizes exceeds 2 to 1, adjustments must be made in γ_x, γ_y, and the notch depth. The solution is given in Section 3, formulas (2.5) and (2.6).

Before developing the details of the solutions, we illustrate the approach on a data set. The example shows how in many practical situations the comparative notches can be used in a multiple comparison of several treatments.

Example

We illustrate the comparative notched box plots on Tippett's (1950) warp break data. Our Figure can be compared to Figure F of McGill et al. (1978). Tippett's data consists of 9 observations each on 6 different types of warp. An observation consists in the number of breaks in a fixed amount of weaving.

A notch depth of 3 determines an exact 82% notch for the population median. From the hypergeometric distribution the two sample comparisons have an exact significance level of 5.7% (see Section 3). Hence from the figure we see that al is significantly greater than bh as judged by a 5.7% Mood two sided test and no other pair yields significant differences at that level.

A 94.3% confidence interval for the difference in population medians (al - bh) is easily found by taking the difference in the notch ends: we find (5, 39).

The quartiles (hinges) occur at depth 3 so the ends of the box coincide with the ends of the notch. We have not drawn in the boxes for this example. The whiskers extend to the farthest observation within one hinge spread of the end of the box. Observations beyond the whiskers are marked by o and should be investigated as possibly stray values. Finally the asymmetry in the notches should be noted since this indicates stretching or compression in the data.

We have not attempted to control the overall error rate for the 15 pairwise comparisons. Using Bonferroni's inequality the overall error rate would be bounded above by 15 X .057 = .855. Since the sample sizes are equal we could set the comparison error rate α_c equal to $\alpha_o/15$ where α_o is the specified overall error rate. Then the notch depths are approximated by (1.5). For example, if $\alpha_o = .15$ so that $\alpha_c = .01$ and $Z_{\alpha_c/2} = 2.576$ we find the depth to be 2 rather than 3 which was used in the example and $\gamma_x = \gamma_y = .93$. The al and bh notches are still disjoint so the comparative statements remain the same.

The approach to multiple comparisons of k samples will work as long as the k(k - 1)/2 ratios of sample sizes do not exceed 2 to 1. For larger ratios the method will not work because more than one notch would be required for each sample.

- Figure -

2. THE APPROXIMATE SOLUTION

We begin with the specified comparison error rate α_c and derive the notch depth formulas and the formula for determining the confidence coefficients γ_x and γ_y.

Suppose γ_x and γ_y, to be determined, are the confidence coefficients for the two notches. When using the approximating distributions we will only consider the case

$$\gamma_x = \gamma_y = \gamma. \tag{2.1}$$

Hence $\alpha = 1 - \gamma$ and the depth d_x are related by (1.3) and similarly for d_y. In case d_x (or d_y) is not an integer taking the depth to be the greatest integer in d_x will produce a slightly wider notch and a slightly conservative confidence coefficient.

Using the same argument as Lehmann (1963, Lemma 4) it is easy to show that the lower end of the X-notch L_x and the upper end of the Y-notch U_y have normal approximating distributions given by

$$\begin{aligned} L_x &= X_{(d_x)} \sim n(\theta_x - \frac{Z_{\alpha/2}}{2\sqrt{n_1}\, f(0)}, \frac{1}{4n_1 f^2(0)}) \\ U_y &= Y_{(n_2-d_y+1)} \sim n(\theta_y + \frac{Z_{\alpha/2}}{2\sqrt{n_2}\, f(0)}, \frac{1}{4n_2 f^2(0)}) \end{aligned} \tag{2.2}$$

where f(0) is the height of the density of F at the median.

One side of the comparative test of H_o: $\Delta = \theta_x - \theta_y = 0$ rejects if $L_x > U_y$. By symmetry of the normal approximating distributions and the independence of the two samples, the two sided significance level is approximately

$$\alpha_c \doteq 2\Phi[-Z_{\alpha/2}(\frac{\sqrt{n_1} + \sqrt{n_2}}{\sqrt{n_1 + n_2}})] \tag{2.3}$$

where $\Phi(\cdot)$ is the standard normal cdf.

Hence for a specified α_c the value of $Z_{\alpha/2}$, needed in (1.3), is given by

$$Z_{\alpha/2} = Z_{\alpha_c/2} \left(\frac{\sqrt{n_1 + n_2}}{\sqrt{n_1} + \sqrt{n_2}} \right) \tag{2.4}$$

and the notch depths are given by

$$\frac{n_i}{2} + .5 - Z_{\alpha_c/2} \frac{\sqrt{n_i}}{2} \left(\frac{\sqrt{n_1 + n_2}}{\sqrt{n_1} + \sqrt{n_2}} \right), \quad i = 1, 2. \tag{2.5}$$

The corresponding value of $\gamma = \gamma_x = \gamma_y$ is then found by using the normal approximation to $P(S < d)$ discussed under (1.2). We have

$$\gamma = 1 - 2\Phi \left[-Z_{\alpha_c/2} \left(\frac{\sqrt{n_1 + n_2}}{\sqrt{n_1} + \sqrt{n_2}} \right) \right] \tag{2.6}$$

Let λ be the ratio of sample sizes and note that

$$\frac{\sqrt{n_1 + n_2}}{\sqrt{n_1} + \sqrt{n_2}} = \frac{\sqrt{1 + \lambda}}{1 + \sqrt{\lambda}} \tag{2.7}$$

The expression in (2.7) varies from .7174 at $\lambda = .5$ to $.7071 \doteq 1/\sqrt{2}$ at $\lambda = 1$. Hence if the ratio of sample sizes is less than 2 to 1, we will use $.71 \doteq 1/\sqrt{2}$ in (2.7) to get (1.5) from (2.5).

When $\alpha_c = .05$, take $Z_{\alpha_c/2} = 2$ and (1.4) follows immediately from (1.5). Furthermore, $\gamma \doteq 1 - 2\Phi(-\sqrt{2}) \doteq .84$ from (2.6).

In summary: If we determine the notch depth from (1.5) with $Z_{\alpha_c/2} = 2$ then we have roughly 84% confidence intervals for the population medians. If we reject the null hypothesis of equal population medians when the notches (confidence intervals) are disjoint then the significance level of this test is roughly 5%. These remarks hold for all but very unbalanced sample sizes in which case (2.6) provides the required confidence coefficient corresponding to depths given by (2.5).

From (2.2) it follows that

$$[Y_{(d_y)} - X_{(n_1-d_x+1)},\ Y_{(n_2-d_y+1)} - X_{(d_x)}] \tag{2.8}$$

is a confidence interval for $\Delta = \theta_y - \theta_x$ with confidence coefficient $\gamma_c = 1 - \alpha_c$ determined in (2.3). Using 84% notches yields an approximate 95% confidence interval for $\Delta = \theta_y - \theta_x$. Hence we find the confidence interval for Δ by taking differences in the ends of the notches in the notched box plot.

The natural point estimate for Δ is simply

$$\hat{\Delta} = \text{med } Y_i - \text{med } X_i, \tag{2.9}$$

the difference in the individual point estimates.

3. THE EXACT SOLUTION

We first discuss Mood's median test for H_0: $\Delta = \theta_y - \theta_x = 0$ vs. H_A: $\Delta \neq 0$. The test is described in detail by Noether (1976,

p. 161). In order to simplify the notation in this section we will replace n_1 by m, the X-sample size and n_2 by n, the Y-sample size. The essential part of the median test is

$$L = \# \, Y_i < M_c \qquad i = 1, 2, \ldots, n \tag{3.1}$$

where M_c is the median of the combined sample. For ease of discussion we will consider the case <u>m+n even</u> so that M_c is the average of the middle two observations in the combined sample. The null hypothesis will be rejected when L is too large or too small. Under H_0: $\Delta = 0$, L has a hypergeometric distribution and the tails of this distribution determine the critical region.

Gastwirth (1968) and Pratt (1964) have pointed out that L can be expressed in the following form:

$$L = \#(Y_{(i)} - X_{(\frac{m+n}{2} - i+1)}) < 0 \quad i = 1, \ldots, n. \tag{3.2}$$

(We will suppose without loss of generality that $m \geq n$.) From this form (which is similar to a one sample sign test form) we immediately have that the Hodges-Lehmann (1963) point estimate of Δ is

$$\hat{\Delta} = \operatorname*{med}_i \, (Y_{(i)} - X_{(\frac{m+n}{2} - i+1)}) \tag{3.3}$$

and the confidence interval for Δ based on L is determined by the d^{th} largest and smallest of the differences in (3.2). Just as in

the case of the one sample sign test, the confidence coefficient γ_c is related to d by

$$P(L < d) = \frac{\alpha_c}{2} \tag{3.4}$$

where $\gamma_c = 1 - \alpha_c$, and L has a hypergeometric distribution.

It is easy to see that the differences in (3.2) are naturally ordered as follows: (recall $m \geq n$)

$$\begin{aligned} Y_{(1)} - X_{(\frac{m+n}{2})} &< Y_{(2)} - X_{(\frac{m+n}{2} - 1)} < \ldots \\ &< Y_{(d)} - X_{(\frac{m+n}{2} - d+1)} < \ldots \\ &< Y_{(n)} - X_{(\frac{m-n}{2} +1)} . \end{aligned} \tag{3.5}$$

This means that $\hat{\Delta}$ in (3.3) becomes:

$$\hat{\Delta} = \underset{i}{\text{med}}\, Y_i - \underset{j}{\text{med}}\, X_j \; , \tag{3.6}$$

which agrees with (2.9). Further when d is defined by (3.4) a $100\gamma_c$% confidence interval for Δ is simply

$$[Y_{(d)} - X_{(\frac{m+n}{2} -d+1)} , \; Y_{(n-d+1)} - X_{(\frac{m-n}{2} +d)}]. \tag{3.7}$$

Finally it should be noted that under H_0: $\Delta = 0$

$$EL = \frac{n}{2} \text{ and } \operatorname{Var} L = \frac{mn}{4(m+n-1)}. \tag{3.8}$$

The normal approximation, with continuity correction, can be applied to yield, from (3.4),

$$d \doteq \frac{n}{2} + .5 - Z_{\alpha_c/2} \sqrt{\frac{mn}{4(m+n-1)}} \quad . \tag{3.9}$$

The two sided size α_c Mood test is equivalent to rejecting H_0: $\Delta = 0$ when 0 is not in the $\gamma_c = 1 - \alpha_c$ confidence interval given by (3.7). We now turn to the relationship between this test or confidence interval and the notches described in (1.2).

We will take apart (3.7) in the obvious way: let $d_y = d$, d determined by (3.4) and let

$$d_x = d_y + \frac{m-n}{2} \tag{3.10}$$

then (3.7) yields the two separate intervals defined by *depths* d_y and d_x. (Compare to (2.8).) The confidence coefficients for the two intervals are given by the binomial distribution discussed under (1.2).

Hence if d is determined exactly by (3.4) or approximately by (3.9) to produce a two sided size α_c Mood test then this Mood test is equivalent to rejecting H_0: $\Delta = 0$ when the notches are disjoint where $d_y = d$ and $d_x = d_y + (m-n)/2$.

Using (3.9) to approximate d and taking $d_y = d$ the Y-confidence coefficient is approximately

$$\gamma_y \doteq 1 - 2\Phi(-Z_{\alpha_c/2} \sqrt{\frac{m}{m+n-1}}) \tag{3.11}$$

while the X-confidence coefficient is approximately

$$\gamma_x \doteq 1 - 2\Phi(-Z_{\alpha_c/2} \sqrt{\frac{n}{m+n-1}}). \qquad (3.12)$$

In case $m = n$, $d_x = d_y$ and (3.11) and (3.12) yield $\gamma_x = \gamma_y \doteq .84$ when $Z_{\alpha_c/2}$ is taken to be 1.96 for a .05 test. This corresponds to (2.7).

For $m \geq n$ a Mood test with level around .05 can be constructed as follows: From a binomial table or Noether's Table E select d_y to yield γ_y at or above .84. Then d_x is determined by (3.10) and will yield γ_x at or below .84. Reject H_0: $\Delta = 0$ if the notches are disjoint. By choosing $\gamma_x \leq .84 \leq \gamma_y$ the level of Mood's test is close to .05. The exact level is found from (3.4) with $d = d_y$.

References

Gastwirth, J. L. (1968), "The First-Median Test: A Two-Sided Version of the Control Median Test," Journal of the American Statistical Association, 63, 692-706.

Hodges, J. L., Jr. and Lehmann, E. L. (1963), "Estimates of Location Based on Rank Tests," Annals of Mathematical Statistics, 34, 598-611.

Hollander, M. and Wolfe, D. A. (1973), Nonparametric Statistical Methods, New York: Wiley.

Lehmann, E. L. (1963), "Nonparametric Confidence Intervals for a Shift Parameter," Annals of Mathematical Statistics, 34, 1507-1512.

_____, (1975), Nonparametrics: Statistical Methods Based on Ranks, San Francisco: Holden-Day.

McGill, Robert, Tukey, John W., and Larsen, Wayne A. (1978), "Variations of Box Plots," The American Statistician, 32, 12-16.

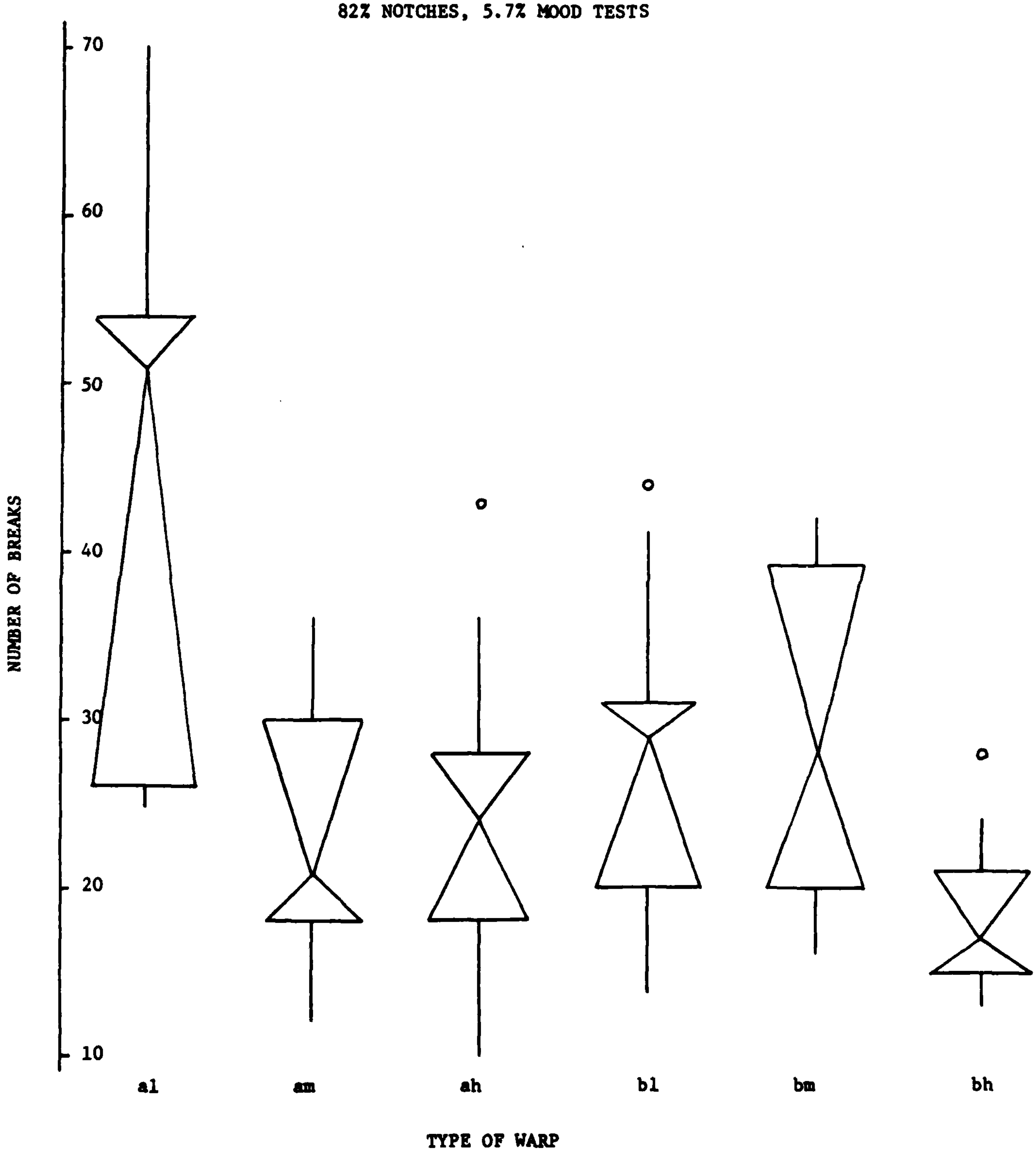
Noether, Gottfried E. (1976), Introduction to Statistics, A Nonparametric Approach, 2nd Edition, Boston: Houghlin Mifflin.

Pratt, John W. (1964), "Robustness of Some Procedures for the Two-Sample Location Problem," Journal of the American Statistical Association, 59, 665-860.

Tippett, L. H. C. (1950), Technological Applications of Statistics, New York: Wiley.

FIGURE

82% NOTCHES, 5.7% MOOD TESTS



Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER
Pennsylvania State University Technical Report #43

2. GOVT ACCESSION NO.
AD-A129 281

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
Comparative Notched Box Plots

5. TYPE OF REPORT & PERIOD COVERED

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)
Thomas P. Hettmansperger, Pennsylvania State University

8. CONTRACT OR GRANT NUMBER(s)
N00014-80-C-0741

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Department of Statistics
The Pennsylvania State University
University Park, PA 16802

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
NR042-446

11. CONTROLLING OFFICE NAME AND ADDRESS
Office of Naval Research
Statistics and Probability Program Code 436
Arlington, VA 22217

12. REPORT DATE
June, 1983

13. NUMBER OF PAGES
14

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

15. SECURITY CLASS. (of this report)
Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

To be submitted to the Journal: Applied Statistics.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Sign test, Mood's median test, Confidence intervals

20. ABSTRACT (Continue on reverse side if necessary and identify by block number) Comparative notched box plots are developed to provide confidence intervals and hypothesis tests for the two sample location model. The notches are confidence intervals derived from the sign test. Rules are given for assigning confidence coefficients to the notches to yield a 95 percent confidence interval and 5 percent two sided test for the difference in locations. The test that rejects no location difference when the "sign" notches are disjoint is shown to be Mood's median test. Circumstances under which multiple comparisons can be carried out are also discussed.